LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-JRNL-235715

A QM/MM Metadynamics Study of the Direct Decarboxylation Mechanism for Orotidine-5'-monophosphate Decarboxylase using Two Different QM Regions: Acceleration too Small to Explain Rate of Enzyme Catalysis

C. Stanton, I-F. W. Kuo, C. J. Mundy, T. Laino, K. N. Houk

October 22, 2007

Journal of Physical Chemistry B

Disclaimer

A QM/MM Metadynamics Study of the Direct Decarboxylation Mechanism for Orotidine-5'-monophosphate Decarboxylase using Two Different QM Regions: Acceleration too Small to Explain Rate of Enzyme Catalysis

Courtney Stanton,[†] I-Feng W. Kuo,[‡] Christopher J. Mundy,[‡§] Teodoro Laino,[¶] K. N. Houk[†]*

Department of Chemistry and Biochemistry, University of California Los Angeles, 607 Charles E. Young Drive East, Los Angeles, CA 90095, Chemical Sciences Division, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, Department of Chemistry and Applied Biosciences ETH Zurich, USI Campus, Via Giuseppe Buffi 13, CH-6900 Lugano, Switzerland

[†] Department of Chemistry and Biochemistry, University of California

[‡] Chemical Science Division, Lawrence Livermore National Laboratory

[§] Current address: Chemical and Materials Sciences Division, Pacific Northwest National Laboratory, Richland, WA 00352

[¶] Department of Chemistry and Applied Biosciences ETH Zurich, USI Campus

[*] Corresponding author e-mail houk@chem.ucla.edu, tel: 310-206-0515, fax: 310-206-1843

Abstract

Despite decades of study, the mechanism by which orotidine-5'-monophosphate decarboxylase (ODCase) catalyzes the decarboxylation of orotidine monophosphate remains unresolved. A computational investigation of the direct decarboxylation mechanism has been performed using mixed quantum mechanical/molecular mechanical (QM/MM) dynamics simulations. The study was performed with the program CP2K that integrates classical dynamics and *ab initio* dynamics based on the Born-Oppenheimer approach. Two different QM regions were explored. The free energy barriers for decarboxylation of orotidine-5'-monophosphate (OMP) in solution and in the enzyme (using the larger QM region) were determined with the metadynamics method to be 40 kcal/mol and 33 kcal/mol, respectively. The calculated change in activation free energy ($\Delta\Delta G^{\pm}$) on going from solution to the enzyme is therefore -7 kcal/mol, far less than the experimental change of -23 kcal/mol (for k_{cat}/k_{uncat} Radzicka, A.; Wolfenden, R., *Science.* **1995,** *267,* 90-92). These results do not support the direct decarboxylation mechanism that has been proposed for the enzyme. However, in the context of QM/MM calculations, it was found that the size of the QM region has a dramatic effect on the calculated reaction barrier.

Keywords: ODCase, proficiency, electrostatic stabilization, *ab initio* dynamics

Introduction

Orotidine-5'-monophosphate decarboxylase (ODCase) was declared the most proficient enzyme known over a decade ago.[1] Since then, there have been many experimental and computational studies attempting to elucidate the mechanism by which this enzyme accelerates the rate of spontaneous decarboxylation of orotidine-5'-monphosphate (OMP) in solution by more than 17 orders of magnitude (Figure 1, a).[1] Surprisingly, the enzyme does not use metal ions or other cofactors, which are commonly essential in other decarboxylases.[2,3,4] Experimental studies have shown that Lys93 (yeast numbering) is important for catalysis but not for binding.[5] Additionally, the ratio of the maximum speed of enzyme reaction, V_{max}, to the dissociation constant for enzyme-substrate complex, K_m, has been shown to peak at pH 7, indicating a catalytic group with a pK_a near 7.[6] There have been several informative reviews on ODCase.[7,8,9]

The direct decarboxylation mechanism for OMP involves stretching of the C6-CO_2 bond, and leads to formation of carbon dioxide and a deprotonated uridine with an unstabilized carbanion at C6 (Figure 1, b). Protonation at C6 either follows or is concerted with decarboxylation to form the final product, uridine monophosphate (UMP). Figure 1, c shows several of the various mechanisms that have been proposed for ODCase, most of which involve stabilization of the negative charge that develops at C6. Decarboxylation of the zwitterionic species formed by protonation at O2 was proposed by Beak and Siegel.[10] Subsequent calculations by Lee and Houk favored decarboxylation accompanied by protonation at O4 rather than O2.[11] This conclusion was further supported by density functional calculations done by Singleton et al. for the reaction in solution.[12] The O4 mechanism was later proposed to include a concerted proton transfer to O4 via a nearby water.[13] Kollman and coworkers suggested that protonation at C5 could catalyze decarboxylation.[14] Michael addition of an active site

nucleophilic residue at C5 was suggested early on by Silverman and Groziak.[15] The importance of an active site lysine led Shostack and Jones to propose Schiff base formation at C4,[16] but lack of ^{18}O exchange of O4 with water seemed to rule that out.

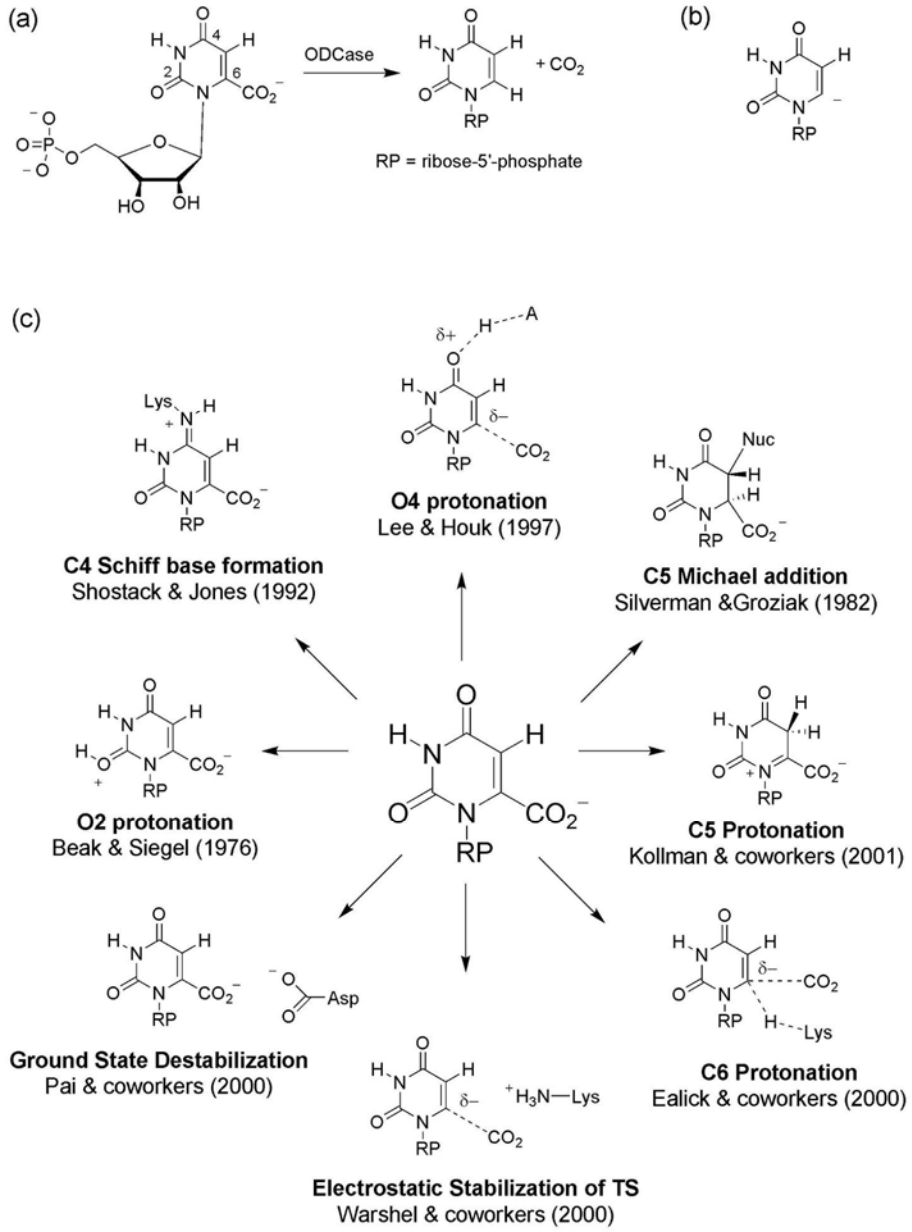


Figure 1. (a) The decarboxylation of OMP to UMP. (b) The carbanion intermediate formed by direct decarboxylation of OMP. (c) An overview of different mechanisms that have been proposed for the reaction in the enzyme. The mechanisms include (clockwise from top) concerted O4 protonation,[11] Michael addition of an active site nucleophile,[15] C5 protonation,[14] concerted C6 protonation/decarboxylation,[17] electrostatic stabilization of transition state,[23] electrostatic stress of the ground state,[20] O2 protonation,[10] Schiff base formation with an active site Lys.[16]

The direct decarboxylation mechanism has gained popularity in recent years due to the lack of experimental evidence for any cofactors or covalent intermediates. Previous QM/MM studies support the direct decarboxylation mechanism, but for different reasons. Gao and coworkers proposed that the proximity of an active-site aspartate to the OMP carboxylate produces a "ground state destabilization" (GSD) effect that facilitates direct decarboxylation.[20] Crystal structures of ODCase from four different species have been reported.[17,18,19,20] In each case, the active site is composed of an array of charged elements. The region expected for the negatively charged carboxylate of OMP is flanked by a Lys-Asp-Lys-Asp quartet, which is generally believed to have each residue in the ionic form. Figure 2 shows the interactions of the substrate OMP in the active site of ODCase after molecular dynamics (MD) equilibration (see later for details). The favorable hydrogen bonding interactions between OMP and the quartet, particularly the ion-pairing of the substrate carboxylate and the nearby protonated lysine, raises serious questions concerning the validity of the GSD hypothesis. The preferential binding affinities for negatively charged inhibitors such as BMP also raise doubt concerning the feasibility of GSD.[21,22]

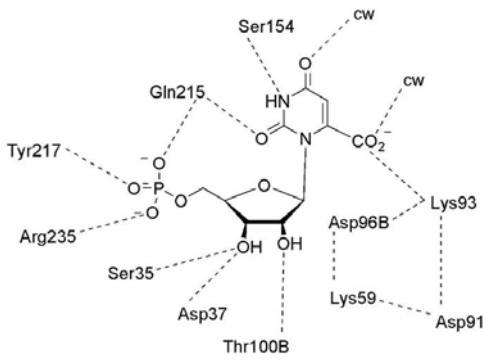


Figure 2. OMP in the active site of ODCase. The dashed lines indicate hydrogen bonding interactions taken from a molecular dynamics simulation (discussed later). cw stands for crystal water. Residue numbering is taken from the yeast enzyme.

A subsequent study by Warshel et al. used an effective valence bond potential (EVB) to investigate the direct decarboxylation of OMP in solution and in the enzyme.[23,24] They propose that the active site is preorganized to maximize favorable interactions with the transition state (TS), and these interactions are sufficient to produce the rate enhancement in the enzyme for direct decarboxylation. Likewise, previous QM/MM studies by Siegbahn and coworkers suggest that direct decarboxylation may be feasible, but they find no evidence for GSD.[25]

A recent study by Carloni and coworkers, which calculated the free energy profile for decarboxylation in the enzyme and in solution, supports a stepwise direct decarboxylation and subsequent C6-protonation mechanism.[26] The authors used a mixed Car-Parrinello *ab initio*/molecular mechanical dynamics method to model ODCase with bound substrate. The free energy profiles were generated using a multiple steering molecular dynamics scheme,[27] one of several nonequilibrium techniques inspired by the work of Jarzynski.[28]

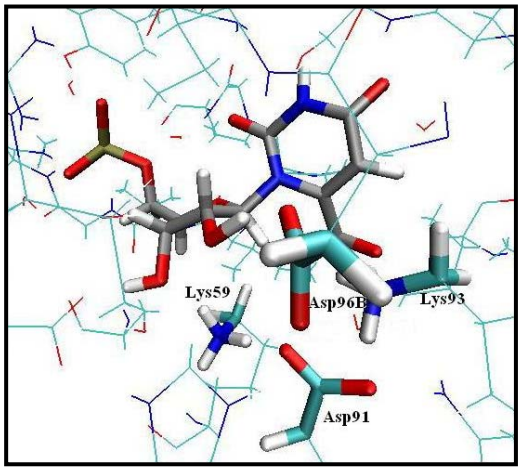


Figure 3. OMP (grey) in the active site of ODCase. The atoms comprising the QM region used in the Carloni and coworkers QM/MM study (Raugei, S.; Cascella, M.; Carloni, P. *J. Am. Chem. Soc*. **2004**, *126*, 15730-15737).

Figure 3 shows the atoms that make up the QM region for these density functional simulations, consisting of the entire substrate OMP and fragments of the side-chains of the Asp-Lys-Asp-Lys quartet as shown in large tubes. They calculated the change in the activation energy for the direct decarboxylation from water to enzyme ($\Delta\Delta G^{\pm}$) to be -22.7 kcal/mol, in good agreement with the experimental $\Delta\Delta G^{\pm}$ of -23.3 kcal/mol.[1]

Appleby and coworkers first proposed a concerted mechanism where C6 of OMP is protonated while the carboxylate bond is broken.[29] This mechanism has been explored in several computational studies. All the studies found the concerted pathway to be higher in energy than a stepwise direct decarboxylation pathway. Warshel et al. used the EVB method to study the direct decarboxylation in solution and the gas phase.[23] They found the concerted pathway to be approximately 20 kcal/mol higher in energy than the direct pathway in solution. In a QM/MM study of the enzymatic reaction, Siegbahn and coworkers calculated the concerted barrier to be

13 kcal/mol higher than direct decarboxylation using QM/MM models of the enzyme.[30] The QM/MM study by Carloni and coworkers also found the concerted process to be higher in energy by about 12 kcal/mol in the enzyme.[26] For these reasons, the concerted pathway was not explicitly studied in this QM/MM study.

The stepwise direct decarboxylation mechanism has been reinvestigated here using a mixed QM/MM method based on Born-Oppenheimer (BO) dynamics, with a much larger QM subsystem (127 atoms) than was used in the previous study by Carloni and coworkers (60 atoms). The free energy profiles for the direct decarboxylation in water and in the enzyme were mapped using metadynamics[31,32] as implemented in the CP2K program,[33] which will be discussed in greater detail in Methods. The results from this study, in contrast to the Carloni work, predict a much larger barrier in the enzyme than the experimental barrier, and therefore do not support a direct decarboxylation mechanism.

Methods

A. QM/MM

All QM/MM calculations were performed using the software suite CP2K, which is freely available at http://cp2k.berlios.de. For the QM/MM simulations, the interaction energy for the QM region was computed via the QuickStep[34] module within CP2K. The QuickStep module performs *O*(N) implementations of density functional theory using a dual basis set method, in which the wavefunctions are described by a Gaussian basis and the density is described by an auxiliary plane wave basis.[34] A triple-ξ Gaussian basis set augmented with two sets of d-type and p-type polarization functions (TZV2P) was used.[35] The plane wave was expanded up to a cutoff of 280 Ry and used in conjunction with the GTH pseudopotential of Goedecker et al.[36,37] to describe the core electrons. Exchange and correlation energy was computed within the GGA

approximation using the BLYP functional.[38,39] For every time step, the electronic structure was explicitly quenched to a tolerance of 10^{-7} Hartree. The interaction energy within the MM subsystem was computed via the FIST module within CP2K.[42] The force field used in this study is the CHARMM force field.[40] The sampling of the potential energy surface was performed using molecular dynamics in the canonical ensemble where every degree of freedom is coupled to one Nose-Hoover chain[41] to ensure thermal equilibration.

The interaction between QM and MM regions was calculated using the procedure described by Laino et al.[42] The MM and QM optimizations are performed separately based on the IMOMM method[43] and electronic embedding. Interactions between QM and MM atoms are not included in the MM calculations, but are included in the QM calculation. In the QM calculation, hydrogen capping atoms are used to fill the empty valences when the QM/MM divide crosses a covalent bond. The electrostatic coupling is calculated using a real space multigrid technique, where each MM atom is represented as a continuous Gaussian function instead of a point charge to avoid the "electron spill-out" problem and to model the QM/MM boundary across a covalent bond more accurately.

B. Metadynamics

Available computer power does not typically allow for adequate sampling in molecular dynamics simulations of large molecules to observe rare events like chemical reactions. Metadynamics is a nonequilibrium method that allows for the system to escape minima in order to sample the rest of the free energy surface on a timescale that is accessible by present day computers.[31,32] The metadynamics method has been used in a number of applications, including the investigation of bacterial chloride channels,[44] deprotonation of formic acid,[45] and flexible ligand docking.[46] The method is based on the assumption that it is possible to define a set of

collective coordinates that can distinguish between reactants and products, and can sample the low-energy reaction paths. Collective variables (CVs) must be functions of the ionic coordinates; examples include bond lengths, dihedral angles, coordination numbers, etc. A history-dependent repulsive potential is built up in low-energy wells by adding a biasing potential term along the CVs at each metadynamic step in the form of a small Gaussian "hill" (equation 2). As the hills build up along the CVs, the system is forced to escape local minima and to explore higher energy regions of the free energy surface (FES). In the limit of infinite time, the biasing potential exactly cancels the underlying FES along the CVs (equation 3):

$$V_{bias}(\mathbf{s},t) = \sum_{t_i} H \exp\left(\frac{\left|\mathbf{s} - s(t_i)\right|^2}{2\omega^2} \right) \quad (2)$$

$$F(\mathbf{s}) = -V_{bias}(\mathbf{s},t)_{t\to\infty} \quad (3)$$

V_{bias} is the repulsive biasing potential term, which is a function of the CVs, **s,** and time, t, with hill parameters having height H and width ω. The FES, $F(\mathbf{s})$, can be reconstructed along the CVs given a sufficient amount of time. An expression for the statistical error associated with equation 3 has been derived[47] (equation 4):

$$\varepsilon = C(d)\sqrt{\frac{S\omega H}{D\tau_G \beta}} \quad (4)$$

C(d) is a constant that depends on the number of CVs, S is the size of the space of interest spanned by the CVs, ω is the hill width, H is the hill height, D is the diffusion coefficient of the collective coordinates in the CV space, τ_G is the metadynamics time step, and β is $(k_B T)^{-1}$.

Computational Details

A. QM model studies

The direct decarboxylation of N-methylorotic acid was first studied using the Gaussian 03 program package,[48] in order to explore the accuracy of BLYP[38,39] for this problem. Density functional calculations were done with both the BLYP and the hybrid B3LYP[49] functionals. The TZV2P basis set, a triple-ξ basis set augmented with two sets of d-type and p-type polarization functions, was used.[35] Continuum solvent calculations were done with the CPCM[50] polarizable continuum model. The high-accuracy CBS-QB3 complete basis set method[51,52] was also used.

B. QM/MM and Metadynamics

For the ODCase/OMP simulations, the entire system used in the QM/MM calculations consisted of the yeast ODCase dimer[53] (PDB code 1DQX) plus the OMP substrate occupying one active site, and the inhibitor BMP occupying the second active site. For the ODCase/BMP simulations, BMP occupied both active sites. Since the active sites on the dimer are known to work independently,[54] only one active site was treated quantum mechanically. Initial coordinates for OMP in the active site of ODCase were determined by overlaying the substrate on one of the bound BMP molecules, and then deleting BMP. The system was then equilibrated with classical dynamics using the molecular dynamics package NAMD[55] for 200 ps with the CHARMM force field[40] before any QM/MM simulations were performed.

Two different QM subsystems were tested in this study. The large QM subsystem was defined as OMP (or BMP) and segments of the following active site residues: Ser35, Asp37, Asp91, Lys59, Asp96B, Lys93, His61, Asn120, Leu153, Ser154, Gln215, Thr100, and two crystal waters (Figure 4). The total number of atoms in the QM region for the ODCase/OMP system was 127 atoms. The size of the QM box was defined as 30 x 18 x 30 Å, while the total QM+MM system including 6821 explicit waters was 60 x 90 x 60 Å. The system was equilibrated with CP2K using QM/MM for an additional 10 ps.

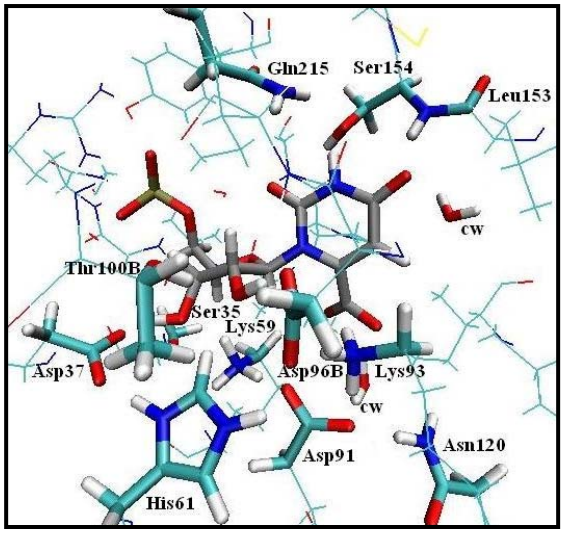


Figure 4. OMP (grey) in the active site of ODCase. The atoms comprising the large QM region used in this study.

For the ODCase/OMP simulations, the CV for the metadynamics simulations was chosen as the distance between C6 and the carbon of the carboxylate substituent of OMP. Several different metadynamics parameters (hill height and width) were tested with classical dynamics to determine reasonable values. Ultimately the parameters chosen for the final metadynamics runs were a height of 1.0 x 10^{-3} Hartree (0.6 kcal/mol) and width of 0.08 Bohr (0.04 Å). Hills were added every 25 MD steps, which had a time step of 0.48 fs. Metadynamics was run for 10 ps in the enzyme after the QM/MM equilibration. To ensure that free diffusivity of the system between the bound and unbound states was reached, the mean square displacement of the $-CO_2$ group along the CV was measured and was found to have a linear dependence on time after approximately 3 ps of metadynamics.

The same method was applied to the small QM subsystem that was used in the Carloni study (Figure 3), including the substrate OMP (or BMP) and fragments of Asp91, Lys59,

Asp96B, and Lys93. The total number of atoms in the QM region for the ODCase/OMP system was 60 atoms. The QM box for the Carloni system was defined here as 18 x 18 x 18 Å, while the size of the total QM+MM system box was the same in both QM/MM simulations, and also contained the same amount of explicit waters.

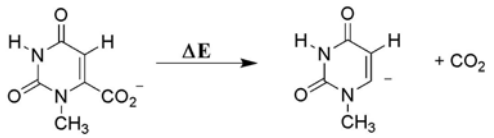
The same parameters were used for both of the enzyme/OMP simulations, and the solution/OMP simulation. For the solution reaction, OMP was surrounded by 23 quantum mechanical waters solvated in a box of 1893 MM waters of dimension 40 x 40 x 40 Å overall. Metadynamics were run for 20 ps in solution after 200 ps of classical equilibration and 10 ps of QM/MM equilibration. Free diffusivity along the CV was reached after 7 ps of metadynamics.

Results

A. QM model studies

Comparison of DFT functionals and basis sets. The BLYP/TZV2P method was used to compute the energetics of the QM subsystem in the QM/MM method described above. Here we compare BLYP/TZV2P to more accurate methods on a smaller system to determine if it gives reasonable results. The two functionals BLYP and B3LYP were compared using both the TZV2P and 6-31+G(d,p) basis sets by calculating the ΔE of the dissociation reaction of carbon dioxide from N-methylorotic acid, as shown in Figure 5 (top). The structures were optimized, and the energy of reaction was calculated for the gas phase decarboxylation shown in Figure 5 (bottom). The hybrid functional B3LYP is generally accepted to be more accurate than BLYP. However, the additional computational cost would be prohibitive within a QM/MM scheme. To gain an estimate of the loss in accuracy by using BLYP instead of B3LYP, the results were compared to the highly accurate complete basis set method CBS-QB3, which has been shown to be accurate

to within 1-2 kcal/mol.[56] The B3LYP/TZV2P combination gave the closest result to the CBS-QB3 method- 34.1 kcal/mol as compared to 35.4 kcal/mol, respectively. The BLYP functional with the TZV2P basis underestimates the CBS-QB3 value by 3.3 kcal/mol. The B3LYP/6-31+G(d,p) method overestimates the ΔE by 2.2 kcal/mol compared to CBS-QB3. The BLYP/TZV2P method used here provides a reasonable estimate, within 3 kcal/mol of the best value available.



Method	Basis Set		
	TZV2P	**6-31+G(d,p)**	
BLYP	32.0		
B3LYP	34.1	37.5	
CBS-QB3			35.3

Figure 5. The computed energies (kcal/mol) of decarboxylation of N-methyl orotate.

B. QM/MM Enzyme Simulations: ODCase/BMP

Crystal structure vs. predicted structure. The crystal structure 1DQX contains the ODCase dimer with bound inhibitor BMP. In order to explore whether the size of the QM subsystem would have an effect on the predicted structure, QM/MM dynamics were performed on the crystal structure using both the small and large QM subsystem as described in Computational Details. Figure 6 shows the results after 4 ps of dynamics. The structures shown for both subsystems were averaged over the last 1 ps of simulation and fitted to the crystal structure with the program VMD[57] using the atoms shown. The RMSD values compared to the crystal structure

for these atoms are 3.1 and 3.4 for the large and small QM subsystem, respectively.

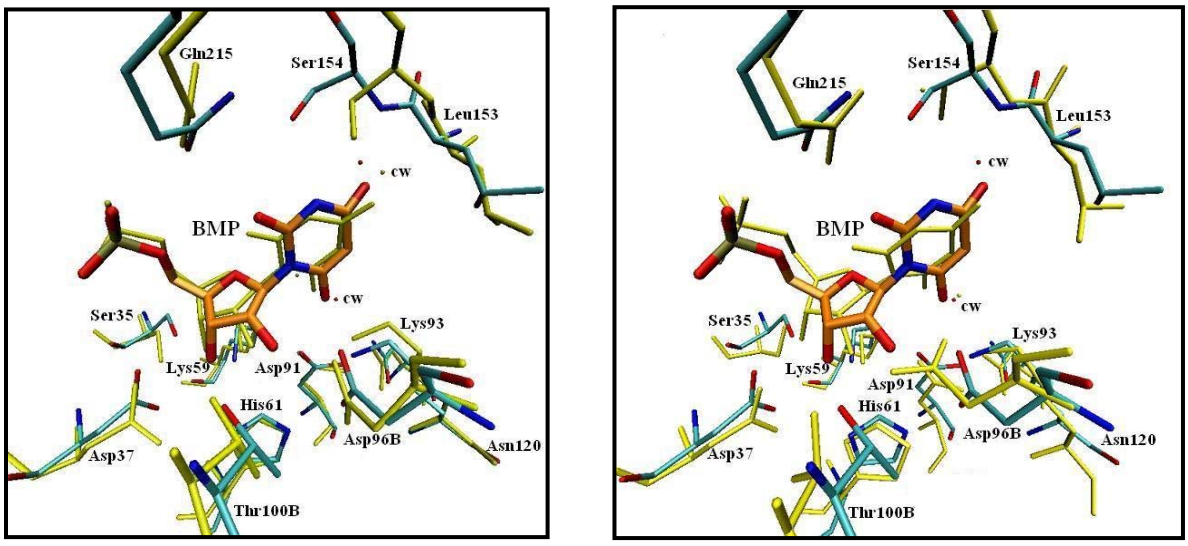


Figure 6. Comparison of crystal structure and CPMD structures, with BMP bound. The crystal structure 1DQX includes bound inhibitor BMP, which is highlighted in orange. The crystal structure is shown in CPK coloring,; cw stands for crystal water. The average structure for each simulation was taken from the last picosecond of a total of 4 ps. (Left) The average structure using the large QM subsystem is shown in yellow. (Right) The average structure using the small QM subsystem is shown in yellow.

In addition to this difference in RMSD, there are significant differences in the protein interactions with BMP as compared to these interactions in the crystal structure.[58] The large QM simulation shows a shorter hydrogen bond between the side chain of Ser154 and O4 of BMP.[59] There are a greater number of deviations from crystal structure for the small QM simulation. A hydrogen bond between Thr100 and 2'-OH of BMP is missing,[60] as is another hydrogen bond between O4 of BMP and a nearby crystal water.[61] The side chain geometry of Asp96B is flipped approximately 90° with respect to the crystal structure, which breaks a hydrogen bond with

nearby His61, and also causes the hydrogen bond with 2'-OH of BMP to be made with the opposite oxygen of the Asp carboxylate. Assuming that the crystal structure is optimal, the small QM simulations appear to introduce substantial deviation from the optimum binding mode for BMP in the active site of ODCase.

C. QM/MM Enzyme Simulations: ODCase/OMP

The large QM subsystem. A single metadynamics simulation was performed for the system with a QM subsystem of 127 atoms, as described in Experimental Details. The collective variable (CV) was defined as the distance between C6 and the carbon of the carboxylate group in the substrate OMP. The free energy as a function of the distance is plotted in Figure 7. At approximately 2 Å, a small bump in the potential corresponds to the breaking of a hydrogen bond between the OMP carboxylate and Lys93. As the bond continues to stretch to about 2.5 Å, another hydrogen bond is broken between the same lysine and an active-site aspartate, while a new hydrogen bond is formed between the lysine and C6 of OMP. The free energy continues to increase until the distance reaches 2.84 Å. This corresponds to the transition state for decarboxylation (Figure 7, **1**), and has a free energy barrier of 33 kcal/mol. The newly formed carbanion is stable for approximately 50 fs of metadynamics until the distance reaches about 3.8 Å. Dynamics generated from a simulation utilizing metadynamics does not correspond to the actual dynamics, so we cannot infer the actual lifetime of the intermediate.[62] At this point in the dynamics, a proton from Lys93 begins to be spontaneously transferred to C6 of OMP. Once the CV has reached about 4.3 Å, the proton has been completely transferred (Figure 7, **2**), forming the final product uridine monophosphate (Figure 7, **3**).

The free energy goes to zero at large values of the collective variable, a characteristic of metadynamics, where there is little to no information about the FES due to lack of hills in that

portion of CV space. The error was calculated according to equation 4, which is a statistical error and is an indication of how well the biasing potential matches the FES in a given portion of CV space. The diffusion coefficient was determined to be 5.0×10^{-9} $m^2 s^{-1}$ by plotting the mean square deviation of the CV versus time. The statistical error associated with this calculation for the CV space between 1.5 and 3.5 Å was calculated to be 1.1 kcal/mol.

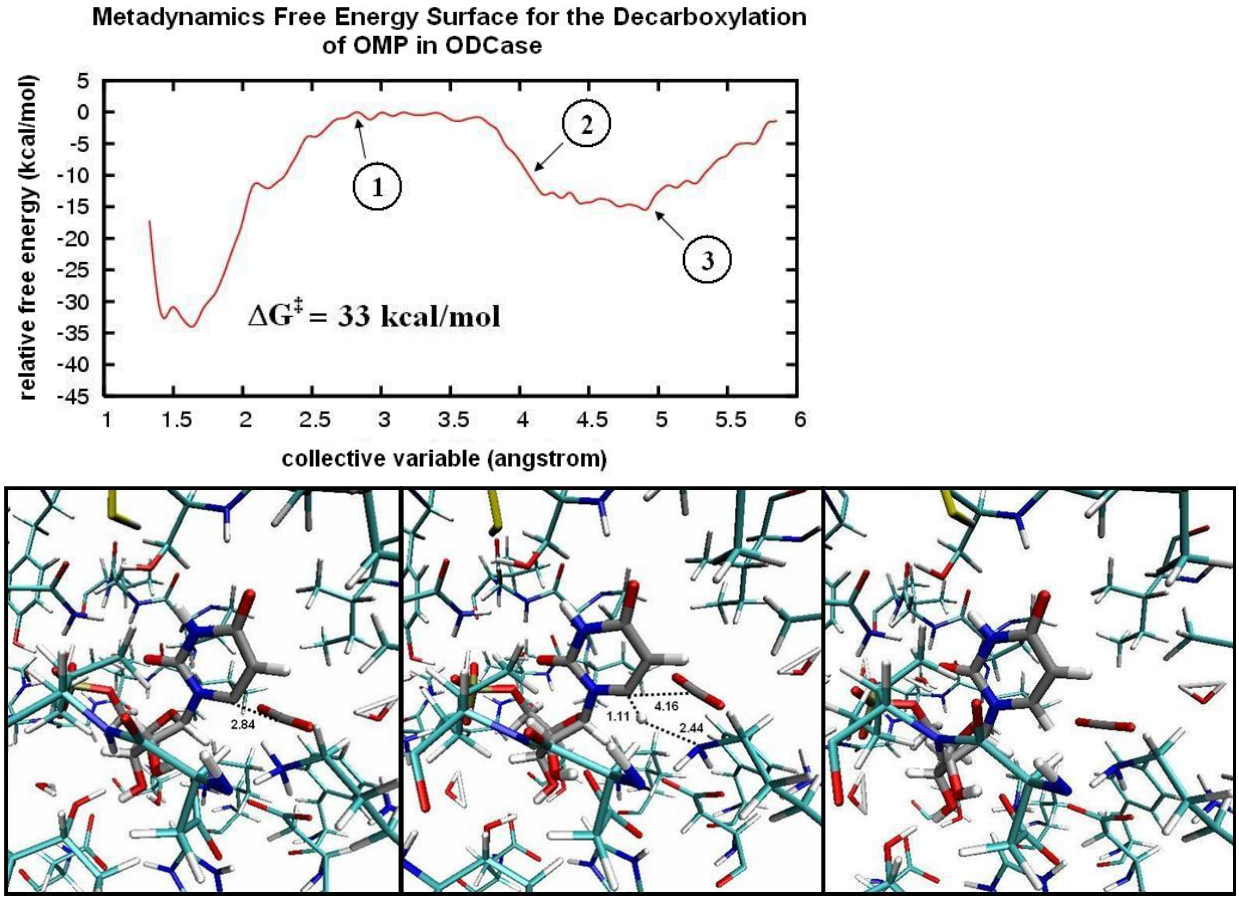


Figure 7. Metadynamics simulation of the direct decarboxylation of OMP in ODCase for the large QM subsystem. (Top) The free energy profile as a function of the CV. The first point marked on the curve, **1**, is the transition structure for decarboxylation. Energies are shown relative to this point. (bottom, left). The next point on the graph, **2**, is after decarboxylation has occurred and a proton is being transferred from an active site lysine to OMP (bottom, middle).The final highlighted point on the graph, **3**, is the stable product (bottom, right). The error associated with the metadynamics was found to be 1.1 kcal/mol for this simulation.

The small QM subsystem. Another metadynamics simulation was run with the smaller QM subsystem used by Carloni et al.,[26] keeping all other parameters and methodology the same as used here for the large QM system. The same hydrogen bond is broken between lysine and the OMP carboxylate when the CV has reached about 2 Å. However, in this case, the free energy reaches a maximum when the CV is only 2.49 Å (Figure 8), which corresponds to a much earlier TS and a smaller reaction barrier. Another important difference is that the carbanion intermediate is less stable in this case, and the protonation at C6 of OMP happens much sooner after intermediate formation. The activation barrier in this case was in good agreement with the Carloni study and calculated to be 24 kcal/mol: 9 kcal/mol less than was calculated for the larger system.

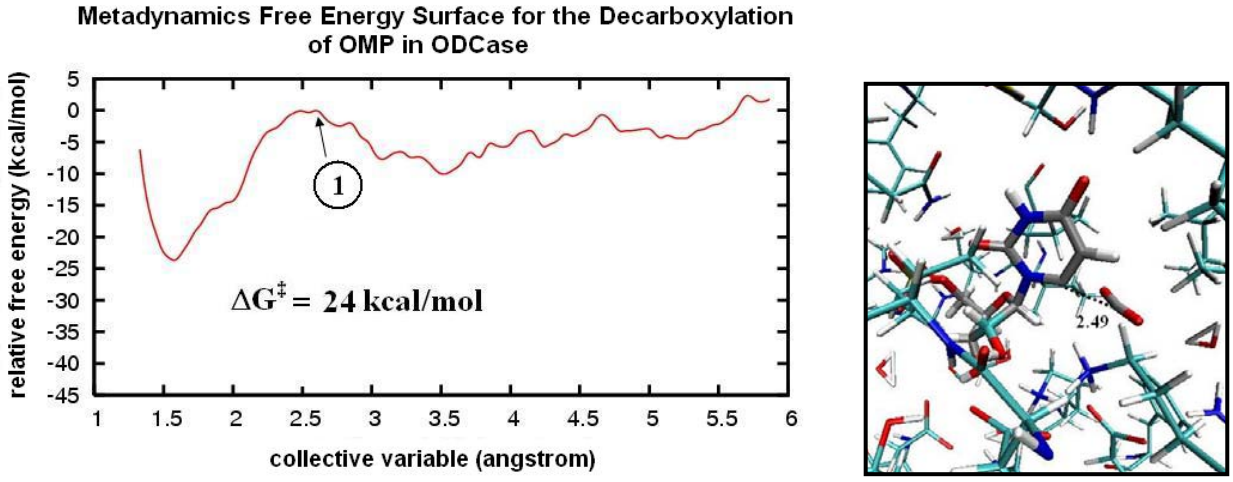


Figure 8. Metadynamics simulation of the direct decarboxylation of OMP in ODCase for the Carloni QM subsystem case. (Left) The free energy profile as a function of the CV. The first point marked on the curve, **1**, is the transition state for decarboxylation. Energies are shown relative to this point. (Right) A snapshot of the point on the graph, **1**, is near the transition state (in the simulation) for decarboxylation.

Small vs. large QM subsystem: how structure explains the difference in reaction barriers. The difference in barrier height for the decarboxylation of OMP in ODCase between the two QM subsystems must be caused by either destabilization of the ground state (GS) or over-stabilization of the TS in the small QM versus the large QM simulations. The various interactions with the GS and TS structures of OMP with the active site residues were examined. Figure 9 shows the interactions that were found to be significantly different (~1 Å or greater) between the two simulations in the GS and in the TS. The ground state is represented on the left-hand side of Figure 9, and the TS is represented on the right-hand side. The numbers in black are the heavy atom distances between OMP and the given residue in the large QM simulation, and the numbers in red are the corresponding distances for the small QM simulation. There are several hydrogen bonding interactions that are diminished in the GS of the small QM system compared to the large QM system. These include the interaction between Ser154 and N3 of OMP, a crystal water (cw) and O4 of OMP, Thr100 and 2'-OH of OMP, and Ser35 and 3'-OH of OMP. The differences in the interactions for the GS were retained in the TS, except for the cw hydrogen bond, which was equalized in the TS.[63] From this analysis, the difference in the activation energy of decarboxylation between the small and large QM system is likely to stem from the unrealistic destabilization of the GS in the active site of the small system.

The residues shown in Figure 9, Ser154, Ser35, and Thr100B, are a part of the large QM region, but not a part of the small QM region. It appears the electrostatic interactions between MM and QM atoms are underestimated compared to between QM-only atoms, giving longer interatomic distances in the small QM region as compared to the large.

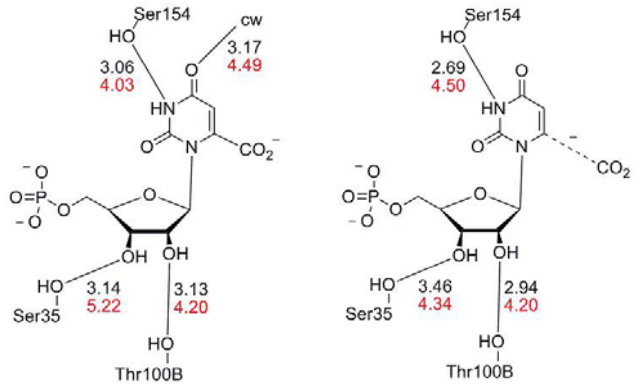


Figure 9. The interactions that were found to be significantly different in the ground state (left) and the transition state (right) for the large and small QM subsystems. The numbers in black are the heavy atom distances from the large QM simulations, and the red numbers are the heavy atom distances from the small QM simulations.

D. QM/MM Solution Simulations: Water/OMP

The same CV and hill parameters were used for the metadynamics simulation in aqueous solution. As the bond stretches, the energy increases until the CV reaches 3.02 Å, giving a free energy barrier to decarboxylation of 40 kcal/mol (Figure 10). A small bump in the FES at about 2.5 Å represents the breaking of a hydrogen bond between C6 and a nearby water molecule. Once reaction occurs, the carbanion intermediate is stabilized with hydrogen bonds from the solvent. Despite the inclusion of 23 QM waters, no proton sharing beyond a typical hydrogen bond was observed. The error associated with this calculation for the CV space between 1.5 and 3.5 Å, and using a diffusion coefficient of 2.5×10^{-9} $m^2 s^{-1}$, is approximately 1.6 kcal/mol. The excellent agreement between prediction and experiment[1] (39 kcal/mol) for the solution reaction is probably somewhat fortuitous, given the inaccuracies in the electronic structure method,

which we found to be approximately 3 kcal/mol.

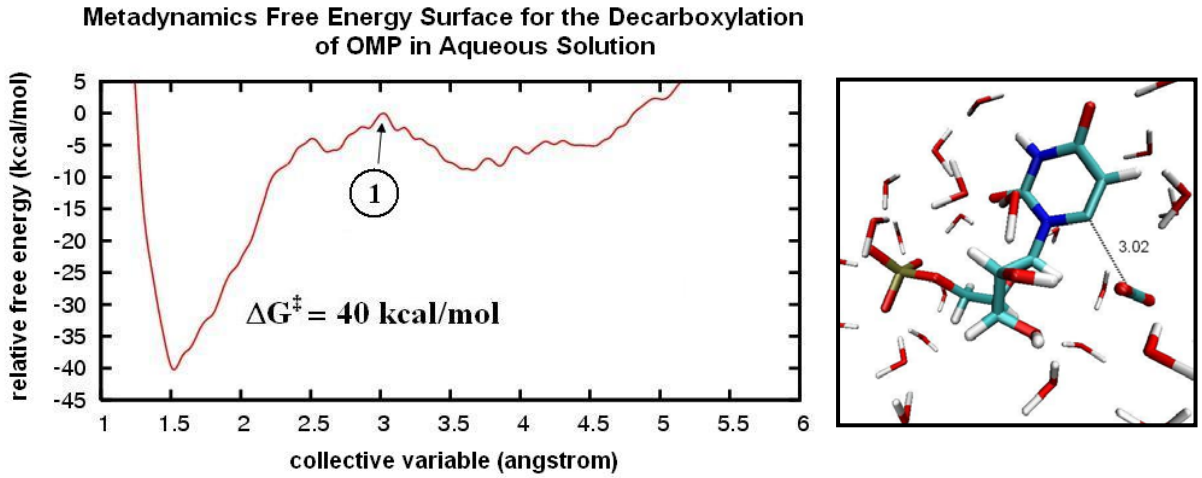


Figure 10. Metadynamics simulation of the direct decarboxylation of OMP in solution. (Left) The free energy profile as a function of CV. The first point marked on the curve, **1**, is the transition state for decarboxylation. Energies are shown relative to this point. (Right) A snapshot of the point on the graph, **1**, is near the transition state (in the simulation) for decarboxylation. The error associated with metadynamics was found to be 1.6 kcal/mol for this simulation.

Discussion

The free energy barrier for the direct decarboxylation mechanism was determined in solution and in the enzyme with the QM/MM metadynamics technique. The enzymatic barrier was calculated to be $\Delta G^{\pm}$ = 33 kcal/mol, while the experimental value is 17 kcal/mol.[1] The calculated solution barrier is closer to the experimental value, $\Delta G^{\pm}$ = 40 kcal/mol versus 39 kcal/mol, respectively. The computed $\Delta\Delta G^{\pm}$ on going from solution to the enzyme is therefore only -7 kcal/mol, much smaller than the experimental value of -23 kcal/mol. The error associated with using the metadynamics technique to calculate free energy barriers was determined to be

1.1 kcal/mol and 1.6 kcal/mol for the enzymatic barrier and the solution barrier, respectively. These results do not support the feasibility of the direct decarboxylation mechanism in the enzyme.

There exists strong circumstantial evidence for chemical catalysis involving a covalent intermediate or general acid-base catalysis.[13,8] A previous study provides a comprehensive survey of hundreds of host-guest complexes and their corresponding kinetic data, including enzymes with transition states.[64] Specifically, it was found that non-covalent complementarity only accounts for catalytic proficiencies[65] of up to 10^{11} M^{-1}, while ODCase exhibits a proficiency of 10^{23} M^{-1}. If ODCase were to catalyze its reaction by non-covalent means, it would be a glaring anomaly among thousands of enzymes**.** This prompted us to reinvestigate the direct decarboxylation mechanism using the most accurate QM/MM methodology applied to the system to date. Improvements over previous QM/MM studies include a much larger QM subsystem, the use of BO dynamics instead of CPMD, and the use of metadynamics, which has a well-defined error function associated with the calculation of free energies. The metadynamics technique is especially attractive since the reactants and products in this case are easily defined by one CV, and unlike other methods, such as umbrella sampling[66] or steered MD[67], the reaction coordinate does not have to be rigorously defined since the method uses a repulsive biasing potential to escape energy minima instead of forcing the system to evolve along a predefined reaction coordinate.

Another metadynamics simulation was performed using the smaller Carloni QM subsystem in order to probe the effect that the size of the defined QM region has on the calculated energetics of the reaction. The size of the QM subsystem was found to have a very significant effect on the dynamics and the energetics. The $\Delta G^{\ddagger}$ of reaction was calculated to be

24 kcal/mol, which is 9 kcal/mol less than the simulation with the larger QM region. The transition state occurs significantly sooner than in the larger QM simulation- at a CV length of 2.5 Å instead of 2.8 Å. The calculated barrier is in good agreement with the 22 kcal/mol barrier that was calculated in the Carloni study. An analysis of the various interactions of OMP with surrounding active site residues in the ground state and the transition state showed that the ground state structure for the small QM subsystem was destabilized compared to the large QM subsystem, resulting in a smaller barrier. There are more MM residues that surround the QM substrate in the small QM simulations compared to the larger QM region, where most of the direct interactions between QM substrate are with QM residues. The interaction between the QM and MM atoms in the small QM simulations is not modeled accurately enough to account for the stabilization seen in the large QM simulations.

This study elucidates the important task of choosing the appropriate QM subsystem in order to capture the crucial interactions due to charge transfer and polarization that are not present in empirical interaction potentials. For the specific case of ODCase, the inclusion of crystal waters as well as residues surrounding the active site give rise to both quantitative and qualitative differences in the observed chemistry.

As mentioned in the Introduction, the direct decarboxylation mechanism has gained support in recent years, including by the first QM/MM study by Pai and Gao.[20] In that study, the QM region was described simply as the reacting orotate ring of the substrate. This gave a barrier of 14.8 kcal/mol, in excellent agreement with experiment (15 kcal/mol). A subsequent EVB study by Warshel et al.[23] utilized a high level *ab initio* derived potential to study the direct decarboxylation reaction of orotate plus ammonium in solution and in the enzyme. They conclude the enzymatic barrier is about 19 kcal/mol, somewhat higher than the previous

calculation. Further QM/MM work by Siegbahn and coworkers found the barrier to direct decarboxylation in the enzyme to be about 22 kcal/mol. The substrate plus the Lys-Asp-Lys-Asp tetrad were treated with the B3LYP/d95 basis set, while the rest of the protein was modeled with the OPLS-AA force field.[68] The recent Carloni and coworkers QM/MM dynamics[26] study using the same QM region as Siegbahn, found the barrier for direct decarboxylation to be 21.5 kcal/mol, in good agreement with the Siegbahn study. It is evident from these studies that the predicted barrier is dependent on the composition of the QM region, and it is not clear that the substrate plus tetrad model is sufficient to obtain accurate reaction energetics. It was confirmed in this study that larger QM regions give a larger calculated barrier height for the direct decarboxylation mechanism.

While some evidence exists that a larger QM region may not be advantageous with semi-empirical QM/MM methods,[69] most QM/MM studies using density functional methods that compare QM region size do not show significant differences in energetics or geometries with increasing QM size.[69,70,71] One recent study using the B3LYP/6-31G(d) QM method with the GROMOS96[72] force field looked at several QM regions to study the planarity of an active site proline in triose phosphate isomerase.[73] It was found that only larger QM regions were able to accurately predict the conformations of this residue.

Additional QM/MM simulations of the ODCase/BMP crystal structure revealed significant structural differences when a larger QM subsystem was defined. The larger QM simulations maintained a lower RMSD with the crystal structure, and several hydrogen bonding interactions with BMP were preserved in the large QM system that were not in the small QM system. These interactions involved several of the residues that were considered QM in the large QM region, but MM in the small QM region simulation. As previously discussed for the

ODCase/OMP simulations, the interaction between QM and MM atoms can be blamed for the discrepancies. This is taken as indirect evidence to support the assumption that a larger QM subsystem gives more accurate results.

These results have broad implications concerning the use of QM/MM methods to study enzymatic reactions and calculating free energies. CP2K uses a sophisticated real space multigrid technique to calculate the QM/MM interaction energy.[42] The MM atomic model itself may be the limiting factor in establishing accurate electrostatic interaction energies between the substrate and the surrounding polar residues. Further calculations are needed to establish the appropriate QM size for a given enzyme/substrate complex. For ODCase/OMP, it seems to be the case that the larger the QM model, the larger the reaction barrier for direct decarboxylation. At this time, calculations using even larger QM regions are prohibitive. Therefore, it is left uncertain whether or not the predicted barrier would continue to increase with a larger QM region.

Conclusion

The direct decarboxylation mechanism of OMP has been investigated by a mixed QM/MM *ab initio* MD study on the entire enzyme/substrate system and in solution. The enzymatic barrier was calculated to be 16 kcal/mol higher than the experimental barrier, while the barrier in solution only deviated by 1 kcal/mol as compared to experiment. The results from the QM/MM study do not support the direct decarboxylation mechanism as the reaction catalyzed by ODCase. However, it was also found that the choice of QM region can have a significant effect on the predicted reaction barrier. A small QM region does not appear to be sufficient to accurately model this reaction in the enzyme, and increasing the size of the QM region tends to increase the calculated barrier.

Explorations of other mechanisms for the decarboxylation of OMP in ODCase with this QM/MM methodology are in progress. Alternative mechanisms that are currently being investigated include Schiff-base formation at C4,[5] concerted O4-protonation/decarboxylation,[11] concerted C5-protonation/decarboxylation,[14] and Michael addition at C5.[15]

Acknowledgement

Funding was provided by the University of California Lawrence Livermore National Laboratory (LLNL) and the National Institutes of Health (NIH). Part of this work was performed under the auspices of the U.S. DOE by University of CA, LLNL under contract No. W-7405-Eng-48. Computer resources were provided by Livermore Computing.

[1] Radzicka, A.; Wolfenden, R., *Science* **1995,** *267,* 90-92.

[2] Steinberger, R.; Westheimer, F. H. *J. Am. Chem. Soc.* **1951,** *73*, 429–435.

[3] Warren, S.; Zerner, B.; Westheimer, F. H.. *Biochemistry* **1966**, *5*, 817–823.

[4] Cleland, W. W. *Acc. Chem. Res.* **1999,** *32*, 862–868.

[5] Shostack, J. A.; Jones, M. E. *Biochemistry* **1992**, *31*, 12162-12168.

[6] Smiley, J. A.; Paneth, M. H.; O'Leary, M. H.; Bell, J. B.; Jones, M. E. *Biochemistry* **1991**, *30*, 6216-6223.

[7] Miller, B. G.; Wolfenden, R. *Ann. Rev. Biochem.* **2002**, *71*, 847-885.

[8] Houk, K. N.; Tantillo, D. J.; Stanton, C.; Hu, Y. *Top. Curr. Chem.* **2004**, *238*, 1-22.

[9] Lee, J. K.; Tantillo, D. J. *Adv. Phys. Org. Chem.* **2003**, *38*, 183-218.

[10] Beak, P.; Siegel, B. *J. Am. Chem. Soc.* **1976**, *98*, 3601–3606.

[11] Lee, J. K.; Houk, K. N. *Science* **1997,** 276, 942–945.

[12] Singleton, D. A.; Merrigan, S. R.; Kim, B. J.; Beak, P.; Phillips, L. M.; Lee, J. K. 2000, J. Am. Chem. Soc., 122, 3296-3300.

[13] Houk, K. N.; Lee, J. K.; Tantillo, D. J.; Bahmanyar, S.; Hietbrink, B. N. *CHEMBIOCHEM* **2001**, *2*, 113-118.

[14] Lee, T. S.; Chong, L.T.; Chodera, J. D.; Kollman, P. *J. Am. Chem. Soc*. **2001**, *123*, 12837-12842.

[15] Silverman, R. B.; Groziak, M. P. *J. Am. Chem. Soc.* **1982**, *104*, 6436–6439.

[16] Shostak, K.; Jones, M. E**.** *Biochemistry* **1992**, *31*, 12155–12161.

[17] Appleby, T. C.; Kinsland, C.; Begley, T. P.; Ealick, S. E. *Proc. Natl. Acad. Sci. USA* **2000,** *97*, 2005-2010.

[18] Miller, B. G.; Hassell, A. M.; Wolfenden, R.; Milburn, M. V.; Short, S. A. *Proc. Natl. Acad. Sci. USA* **2000,** *97*, 2011-2016.

[19] Harris, P.; Poulsen, J. -C. N.; Jense, K. F.; Larsen, S. *Biochemistry* **2000**, *39,* 4217-4224.

[20] Wu, N.; Mo, Y.; Gao, J.; Pai, E. F. *Proc. Natl. Acad. Sci. USA* **2000***, 97*, 2017-2022.

[21] Levine, H. L.; Brody, R. S.; Westheimer, F. H. *Biochemistry* **1980,** *19*, 4993–4999.

[22] Miller, B. G.; Wolfenden, R. *Annu. Rev. Biochem*. **2002**, *71*, 847-885.

[23] Warshel, A.; Štrajbl, M.; Villà, J.; Florián, J *Biochemistry,* **2000,** *39*, 14728–14738.

[24] Warshel, A.; Florián, J.; Štrajbl, M.; Villà, J. *ChemBioChem* **2001,** *2*, 109–111.

[25] Lundberg, M.; Blomberg, M. R. A.; Siegbahn, P. E. M. *Top. Curr. Chem*. **2004**, *238*, 79-112.

[26] Raugei, S.; Cascella, M.; Carloni, P. *J. Am. Chem. Soc*. **2004**, *126*, 15730-15737.

[27] Cascella, M.; Guidoni, L.; Rothlisberger, U.; Maritan, A.; Carloni, P. *J. Phys. Chem. B* **2002**, *106*, 13207-13032.

[28] Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690-2693.

[29] Appleby, T. C.; Kinsland, C.; Begley, T. P.; Ealick, S. E. *PNAS* **2000**, *97*, 2005-2010.

[30] Lundberg, M.; Blomberg, M. R. A.; Siegbahn, P. E. M. *Top. Curr. Chem.* **2004**, *238*, 79-112.

[31] Laio, A.; Parrinello, M. *PNAS* **2002**, *99*, 12562-12566.

[32] Micheletti, C.; Laio, A.; Parrinello, M. *Phys. Rev. Lett*. **2004**, *92*, 170601.

[33] See http://cp2k.berlios.de/ for more information on CP2K

[34] VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comp. Phys. Com.* **2005**, *167*, 103-128.

[35] Schaefer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829.

[36] Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703.

[37] Hartwigsen, C.; Goedecker, S.; Hutter, *J. Phys. Rev. B* **1998**, *58*, 3641.

[38] Becke, A. D. *Phys. Rev. A* **1988,** *38*, 3098.

[39] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

[40] MacKerell, Jr. A. D.; Brooks, B.; Brooks, III, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus M. *The Encyclopedia of Computational Chemistry*, **1**, 271-277, P. v. R. Schleyer et al., editors (John Wiley & Sons: Chichester, 1998)

[41] Nosé, S. *J. Chem. Phys*. **1984**, *81*, 511-519.

[42] Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. *J. Chem. Theory Comput.* **2005**, *1*, 1176-1184.

[43] Maseras, F.; Morokuma, K. *J. Comp. Chem.* **1995**, *16*, 1170-1179.

[44] Gervasio, F. L.; Parrinello, M.; Ceccarelli, M.; Klein, M. L. *J. Mol. Biol.* **2006**, *361*, 390-398.

[45] Lee, J. G.; Asciutto, E.; Babin, V.; Sagui, C.; Darden, T.; Roland, C. *J. Phys. Chem. B* **2006**, *110*, 2325-2331.

[46] Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc*. **2005**, *127*, 2600-2607.

[47] Laio, A.; Rodriguez-Fortes, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. *J. Phys. Chem. B* **2005**, *109*, 6714-6721.

[48] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.;

Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision D.01; Gaussian, Inc.: Wallingford, CT, 2004.

[49] Becke, A. D. *J. Chem. Phys.* **1993***, 98*, 5648.

[50] Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998,** *102*, 1995.

[51] Montgomery Jr., J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **2000**, *112*, 6532.

[52] Montgomery Jr, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, *110*, 2822.

[53] Yablonski, M. J.; Pasek, D. A.; Han, B. D.; Jones, M. E.; Traut, T. W. *J. Biol.Chem.,* **1996,** *271*, 10704–8

[54] Porter, D. J. T.; Short, S. A. **2000**, *Biochemistry*, *3*9, 11788–800

[55] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comp. Chem*. **2005**, *26*, 1781-1802.

[56] Pokon, E. K.; Liptak, M. D.; Feldgus, S.; Shields, G. C. *J. Phys. Chem. A* **2001**, *105*, 10483-10487.

[57] Humphrey, W.; Dalke, A.; Schulten, K. *J. Molec. Graphics* **1996**, *14*, 33-38.

[58] In both simulations, the side chain of Gln215 is flipped 180° from the crystal structre conformation so that $-NH_2$ of the amide side chain forms hydrogen bonds with the phosphate group and O2 of BMP. The reference for the crystal structure (ref. 13), describes a hydrogen bond between Gln215 and BMP at O2. It is likely there is a mistake in the PDB file and the positions of the N and O of the amide should be reversed.

[59] The heavy atom distances are 3.76 and 2.66 Å in the crystal structure and simulation, respectively.

[60] The heavy atom distances are 2.76 and 4.20 Å in the crystal structure and simulation, respectively.

[61] The heavy atom distances are 3.17 and 5.74 Å in the crystal structure and simulation, respectively.

[62] Ensing, B.; Vivo, M. D.; Liu, Z.; Moore, P.; Klein, M. *Acc. Chem. Res.* **2006**, *39*, 73-81.

[63] The heavy atom distances are 2.92 and 3.20 Å in the large and small QM TS structures, respectively.

[64] Zhang, X.; Houk, K. N. *Acc. Chem. Res.* **2005**, *38*, 379-385.

[65] Catalytic proficiency is defined as: $\frac{k_{cat} / K_m}{k_{uncat}}$

[66] Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic: San Diego, CA 2002.

[67] Gullingsrud, J. R.; Braun, R.; Chulten, K. *J. Comput. Phys.* **1999**, *151*, 190-211.

[68] Jorgensen, W. I.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996***, 118*, 11225-11236.

[69] Lennartz, C.; Schäfer, A.; Terstegen, F.; Thiel, W. *J Phys Chem B* **2002,** *106*,1758–1767.

[70] Loeffer, H. H.; Mohammed, A. M.; Inada, Y.; Funahashi, S. *Chem. Phys. Lett.* **2003**, *379*, 452-457.

[71] Bathelt, C. M.; Zurek, J.; Mulholland, A. J.; Harvey, J. N. *JACS* **2005**, *127*, 12900-12908.

[72] van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G.; Biomolecular simulation: GROMOS96 manual and user guide. Zürich, Groningen: BIOMOS b.v; 1996

[73] Donnini, S.; Groenhof, G.; Wierenga, R. K.; Juffer, A. H. Prot.*: Struc., Funct., Bioinf.* **2006**, *64*, 700-710.